Virtual Center for Network and Security Data

University of Michigan
Merit Network
University of Wisconsin
Georgia Tech
University of Washington

PREDICT PI MTG

Monday, March 21, 2011

University of California San Diego

Virtual Center for Network and Security Data

- Virtual Center for Network and Security Data is a collaborative effort among four institutions with the University of Michigan as lead. Its goals are to:
  - Develop a virtual data repository of rich, correlated, security datasets representing Internet scale behaviors
  - Provide a wide range of extremely relevant data sources, providing a global perspective on Internet behavior
  - Providing distributed services to simplify accessing and using this data

Participants

- Funding Phase I
  - Farnam Jahanian, University of Michigan
  - Morley Mao, BEACON
  - Manish Karir, Merit Network
- Funding Phase II
  - Paul Barford, University of Wisconsin
  - Dave Dittrich, University of Washington
  - Matthew Zekauskas and Rick Summerhill, Internet2
- Funding Phase III
  - Paul Barford, University of Wisconsin
  - Nick Feamster, Georgia Tech
- Menlo and Dataset Subcontracts
  - University of Washington
  - Wenke Lee, Georgia Tech

Merit Network, Inc. - PREDICT Dashboard

System Status | System Log | Dataset Descriptions | Reading Library | Search | About

Current System Status(UTC):

Netflow:	Up since February 1, 2010, 8:03 am
Darknet:	Up since February 1, 2010, 6:00 am
BGP:	Up since February 2, 2010, 9:13 am
DNS:	Up since July 23, 2010, 3:58 pm

Data Collection Statistics:

Overall Cumulative Stats

Overall Total Stored Data:	**122.29T**
Overall Total No of Files:	312,455
Overall Total Free Space:	19.03T
Yesterday Data Growth:	98.64G

Netflow Details

Total Stored Data:	32.88T
Total No of Files:	231,113
Total Free Space:	8.67T
Yesterday Data Growth:	41.50G

Darknet Details

Total Stored Data:	88.31T
Total No of Files:	47,283
Total Free Space:	7.16T
Yesterday Data Growth:	54.37G

BGP Details

Total Stored Data:	845.01G
Total No of Files:	33,602
Total Free Space:	3.20T (shared with DNS)
Yesterday Data Growth:	1.42G

DNS Details

Total Stored Data:	290.11G
Total No of Files:	457
Total Free Space:	3.20T (shared with BGP)
Yesterday Data Growth:	1.35G

1000 Oakbrook Drive, Suite 200, Ann Arbor, Michigan 48104-6794
Phone: (734) 527-5700 Fax: (734) 527-5790 E-mail: www@merit.edu

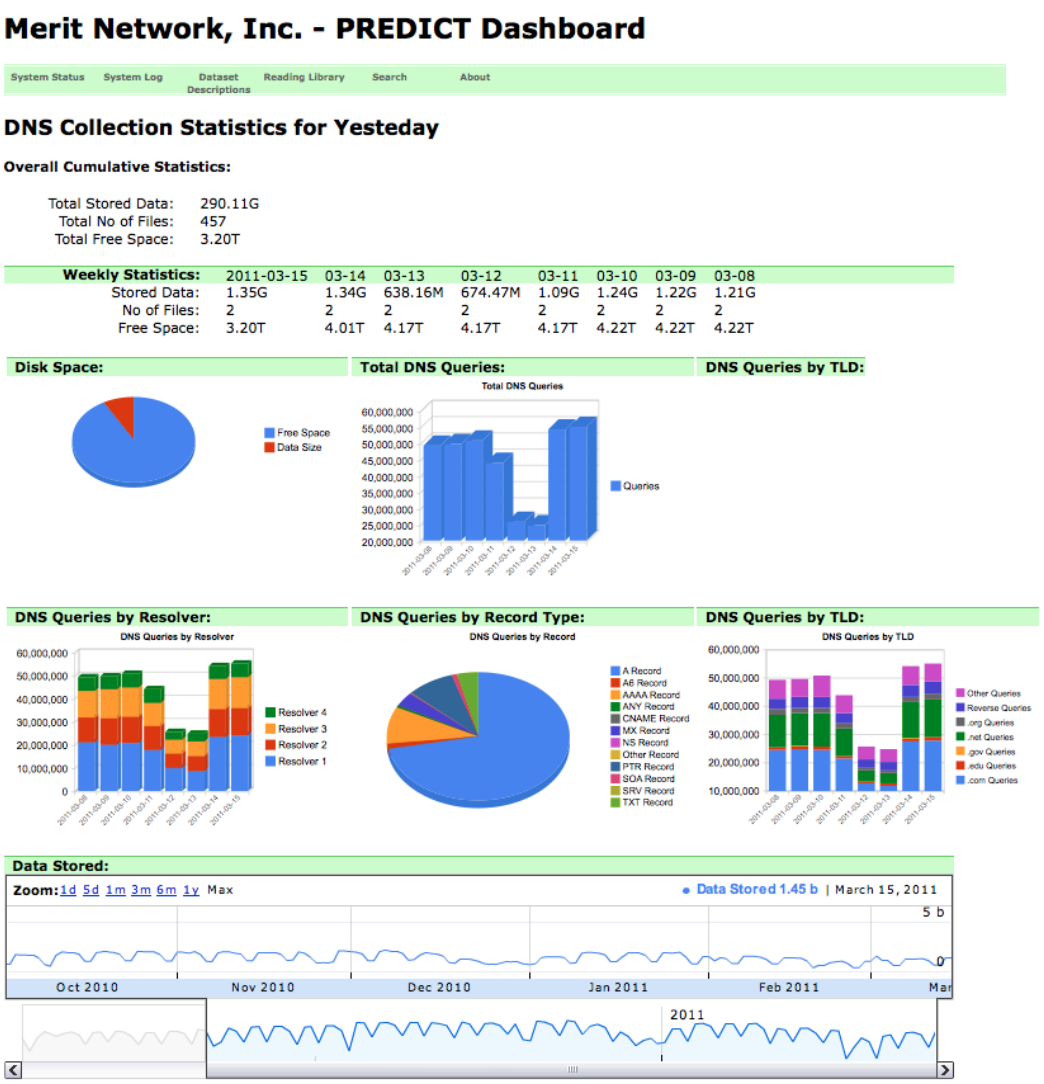
Dataset Statistics

Dataset	This Period*
ims-bagle_backdoor	2
ims-dabber	2
ims-mydoom_backdoor	2
ims-mysqlbot	2
ims-sasser	2
ims-tcp42_wins	2
ims-tcp6101_veritas	2
ims-trend-1	2
ims-witty	2

* Issue #82 has been updated by Kyle Creyts.

manish, i still haven't heard anything from the fellow from clarkson. I will forward you another copy of the email in a few minutes.

New DNS Data Collection



- DNS dataset now incorporated into our monitoring/ reporting structure
- Roughly 1-1.3GB data/day – 6 months
- 4 Resolvers – queries only no responses
- Sources of queries anonymized by hashing (SHA1) on a per day basis
- 25M-50M queries per day

DNS Data

- The Internet Domain Name System (DNS) is a distributed hierarchical naming system that at its most fundamental level provides a mapping between IP address and names. Examples of data in the DNS data category include DNS traffic (queries and/or responses), DNS server logs, and other DNS related meta data. These data may be collected at or near clients, recursive resolvers, or DNS servers for an enterprise, top-level, or root domain. Contents employ appropriate mechanisms to protect privacy. Possible uses of data in the DNS data category are: studying the DNS performance, detecting malicious behavior on the Internet (i.e., fast flux domain-based botnet detection), estimating the amount of network activity, inferring relative popularity of Internet applications.

Dataset creation via annotation

- Use Arbor Peakflow Alerts
- Attempt to validate event with ground truth (manual checks on live flows if event is still active, noc/ops engineer conversations)
- Extract netflow subset that includes the interesting traffic

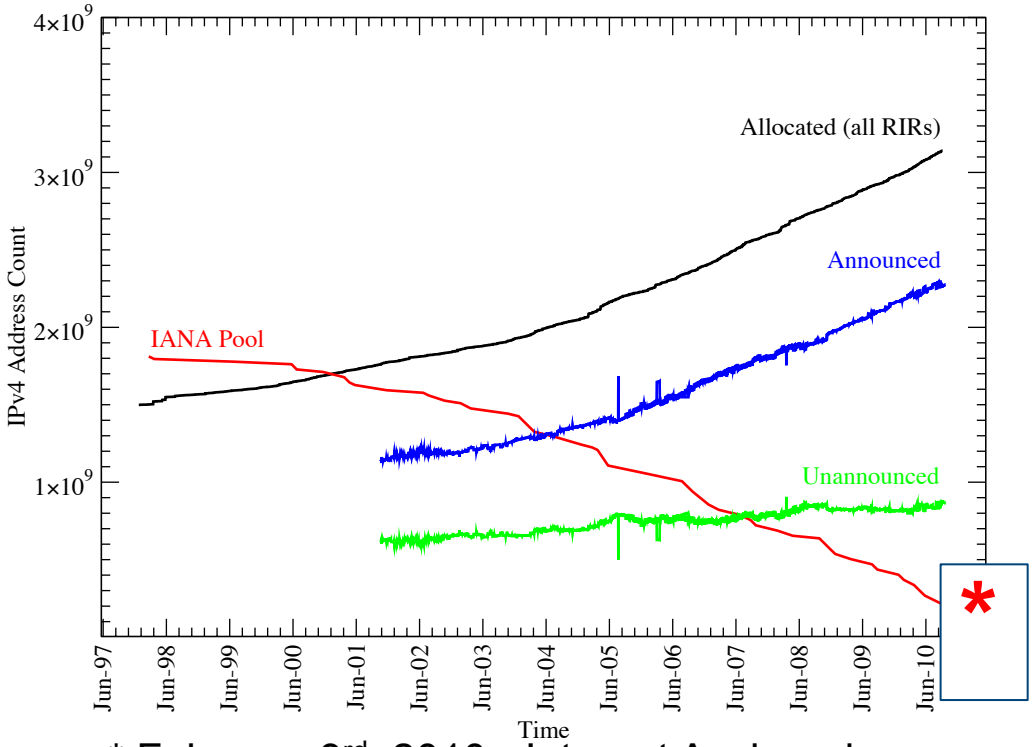
New DDoS Dataset – syn-flood attack

- Goal: To create annotated sub-datasets from larger netflow data collection:
- syn-flood attack netflow dataset
  - syn-flood attack directed at UM IRC server
  - 10min duration (2011-03-04 16:18:01 to 2011-03-04 16:28:18)
  - Fingerprint: `(proto 6 and (src port 3072 or src port 1024) and dst port 6667) and src 0.0.0.0/0 and dst 141.213.238.252/32`
  - Peak traffic volume 139Mbps (362Kpps)
  - File size: 21GB (20 min subset of all merit netflow data)
  - Dataset Duration: 2011-03-04 16:08:00 to 2011-03-04 16:30:25
  - Anonymized

New Darknet Datasets

- Complete set Darknet packet capture samples of last 26 /8 allocations for IPv4
- 13 additional captures since last update (15TB)
- Total set: 1/8, 5/8, 14/8, 23/8, 31/8, 36/8, 37/8, 39/8, 42/8, 45/8, 49/8, 50/8, 100/8, *101/8, 102/8, 103/8, 104/8, 105/8, 106/8, 107/8, 176/8, 177/8, 179/8, 181/8, 184/8, 223/8*

The Fall of the Today's Internet



* February 3rd, 2010 - Internet Assigned Numbers Authority (IANA) assigns the last of its IPv4 addresses

"Within the next 12 to 18 months, or perhaps sooner, every one of the 4.3 billion Internet Protocol addresses will have been allocated, and the Internet, at least as it exists today, will have reached full capacity. " NY Times

End of IPv4?

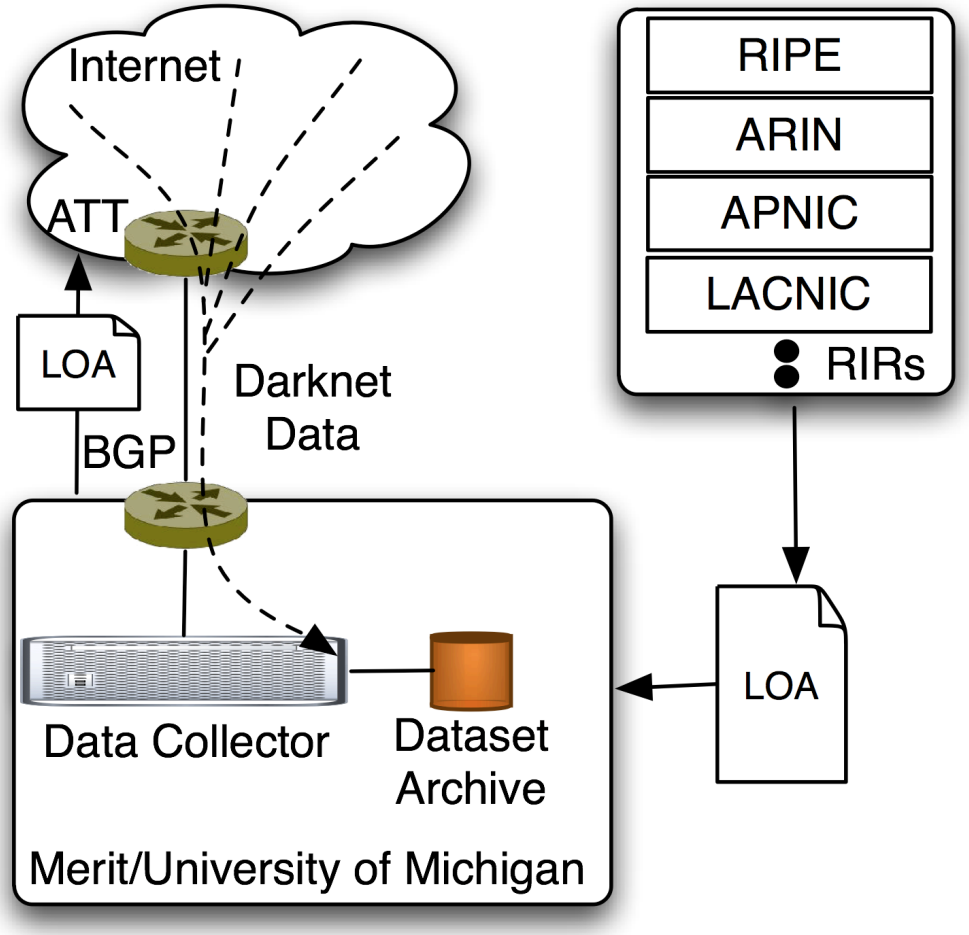
- Monitor the “last” of IPv4 space before the RIRs start allocating.
  - 102/8 AFNIC
  - 103/8 APNIC
  - 179/8 LACNIC
  - 104/8 ARIN
  - 185/8 RIPE

Are we Done then?

- More analysis!!!!
- Mobile Darknets
  - Identify mobile space, look at mobile scanning behavior
  - Darknets in Mobile space
- IPv6 Darknets
  - /14 with LACNIC waiting on LOA
- ERX Recaptured Networks
  - LACNIC wants to capture data on 30 /16s being returned

Internet Pollution – Redux

A Framework for Internet Pollution Analysis

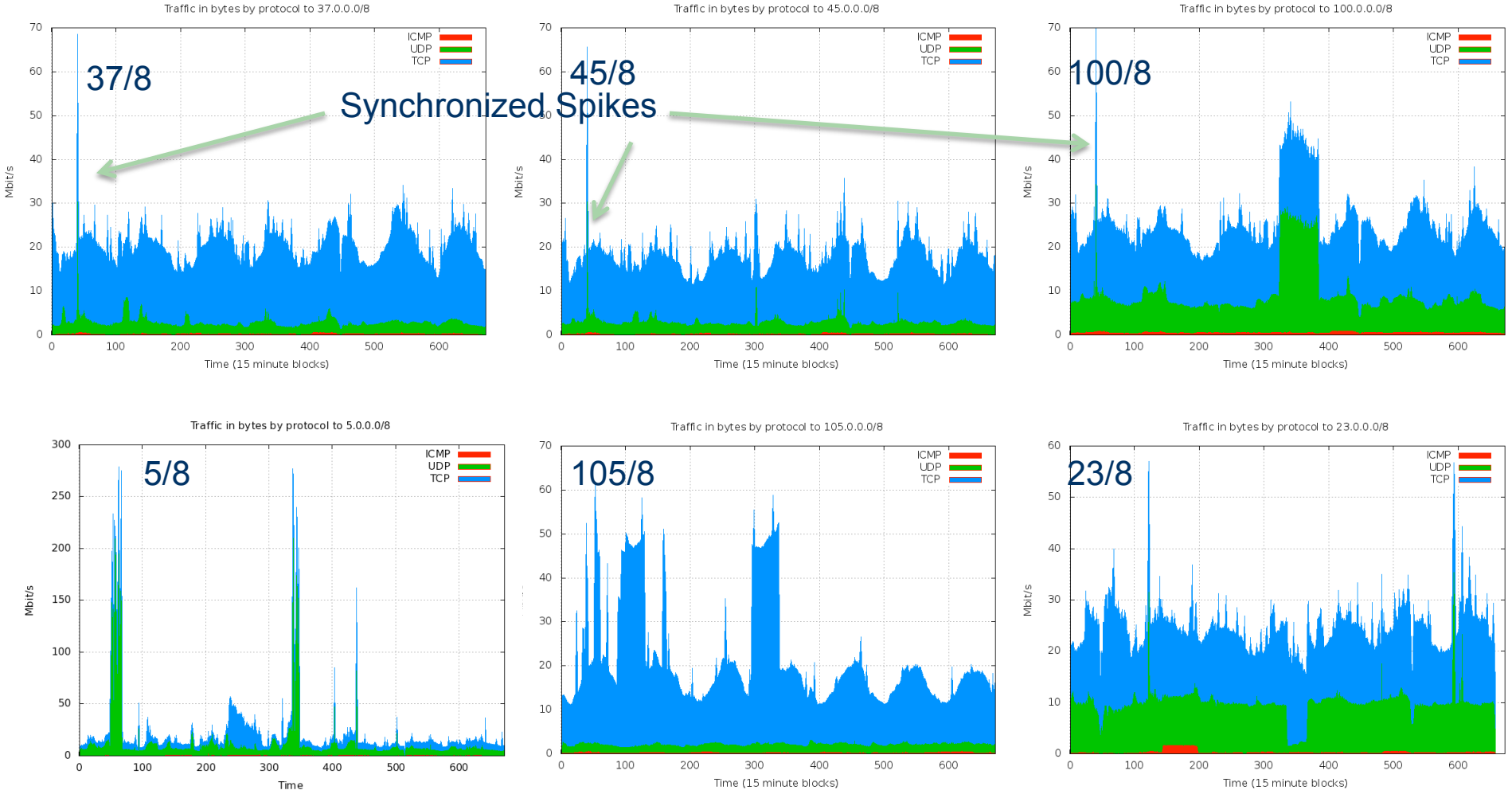


- Work with RIRs to identify upcoming allocation
- Obtain LOA
- Advertise, Collect, Analyze, Archive, Provide to research community
- Cleanup/Quarantine recommendations
- Support from DHS via PREDICT Project

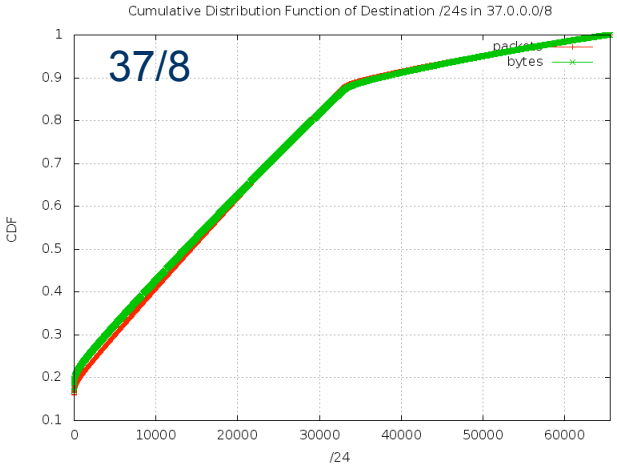
Cross RIR Darknet Traffic Analysis

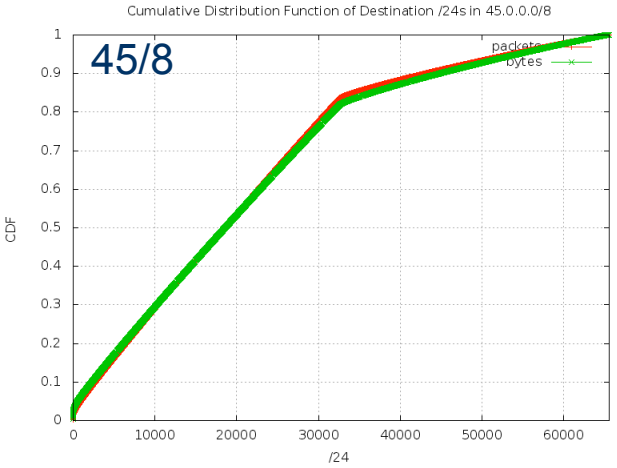
- Goal: Analyze darknet traffic to determine how much and what kinds of pollution were present in each block and determine whether cleanup/quarantine were viable options
- 23/8, 100/8, 45/8 - ARIN
- 5/8, 37/8 - RIPE
- 105/8 – AfriNIC
- Several 7 day long datasets were collected – here we are presenting results from a 6 /8 collection with 3 simultaneous announcements (37, 45, 100)
- Alternate dataset has all 6 /8 announcements at the same time

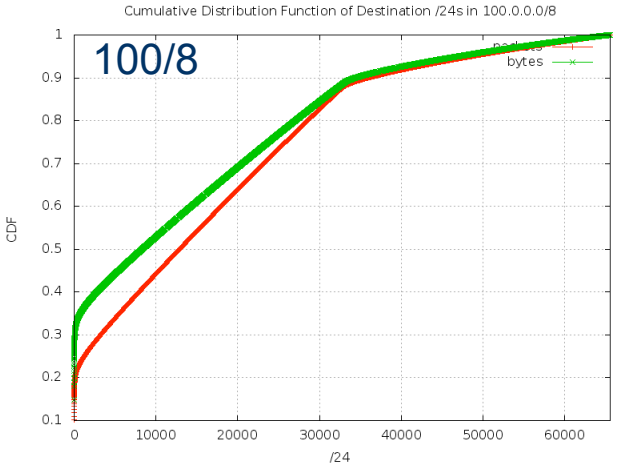
Comparing Traffic Volumes

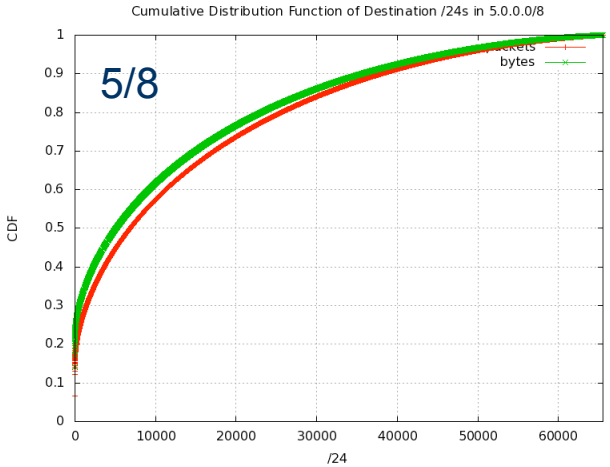


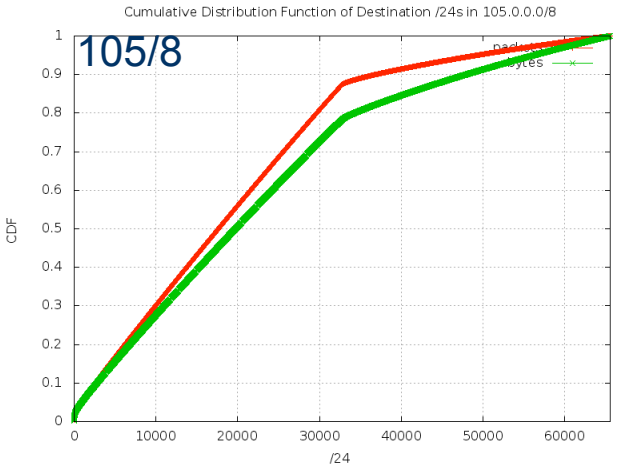
Comparing Hotspot Activity

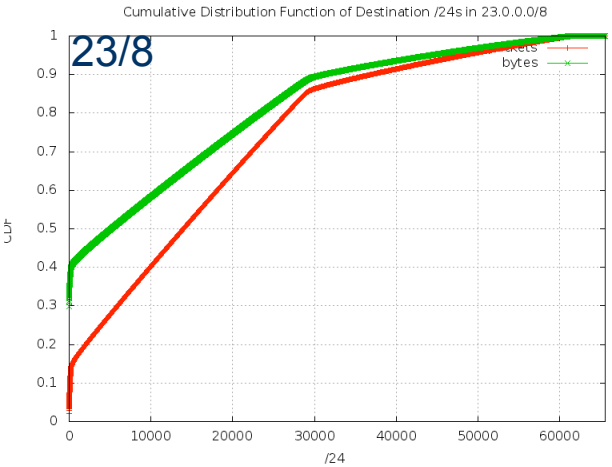




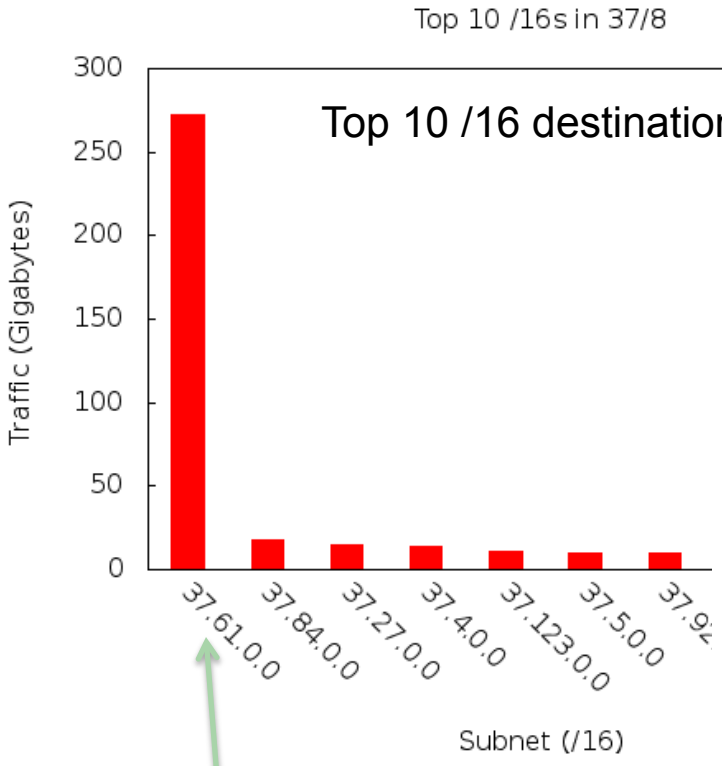


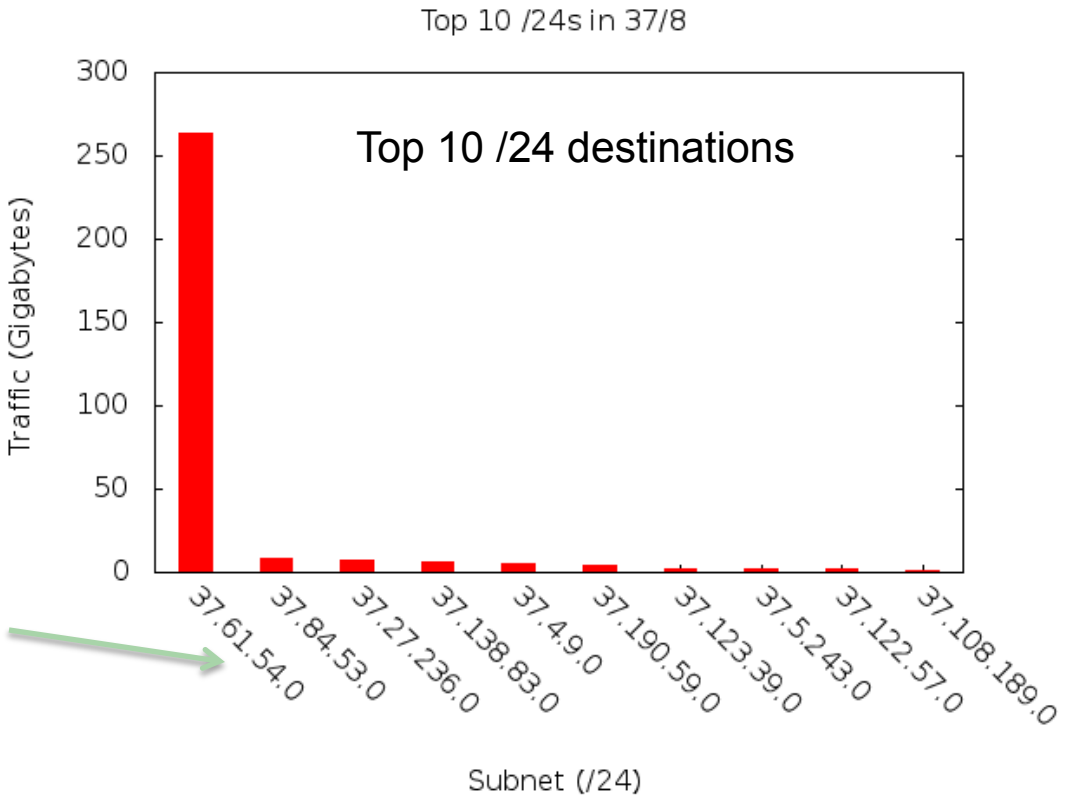






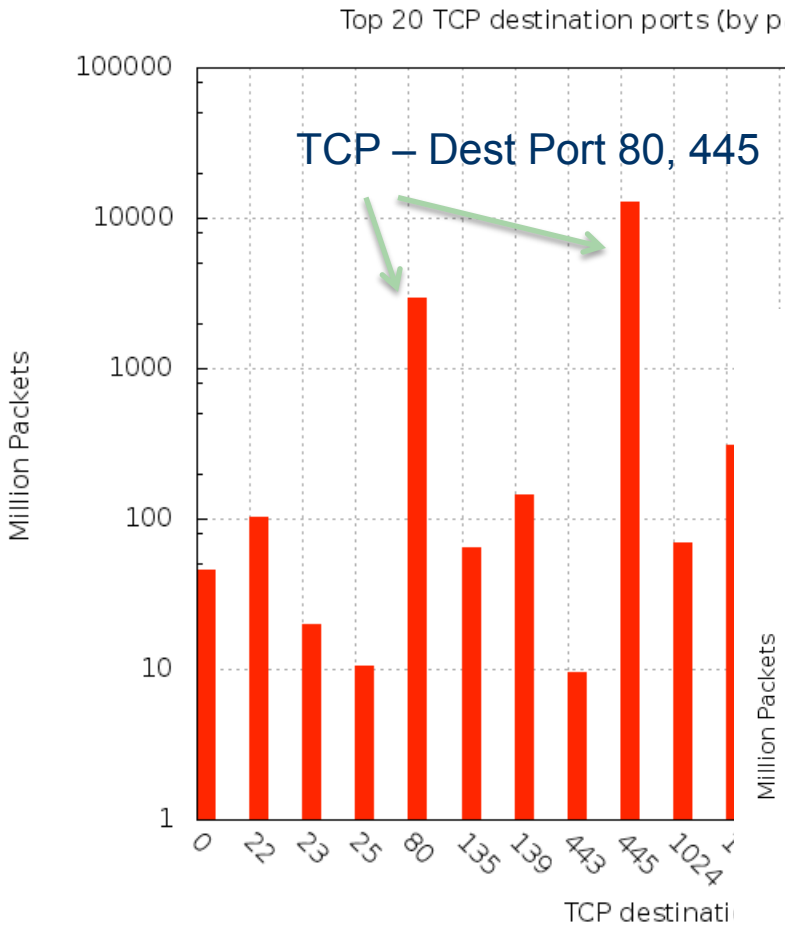
37/8

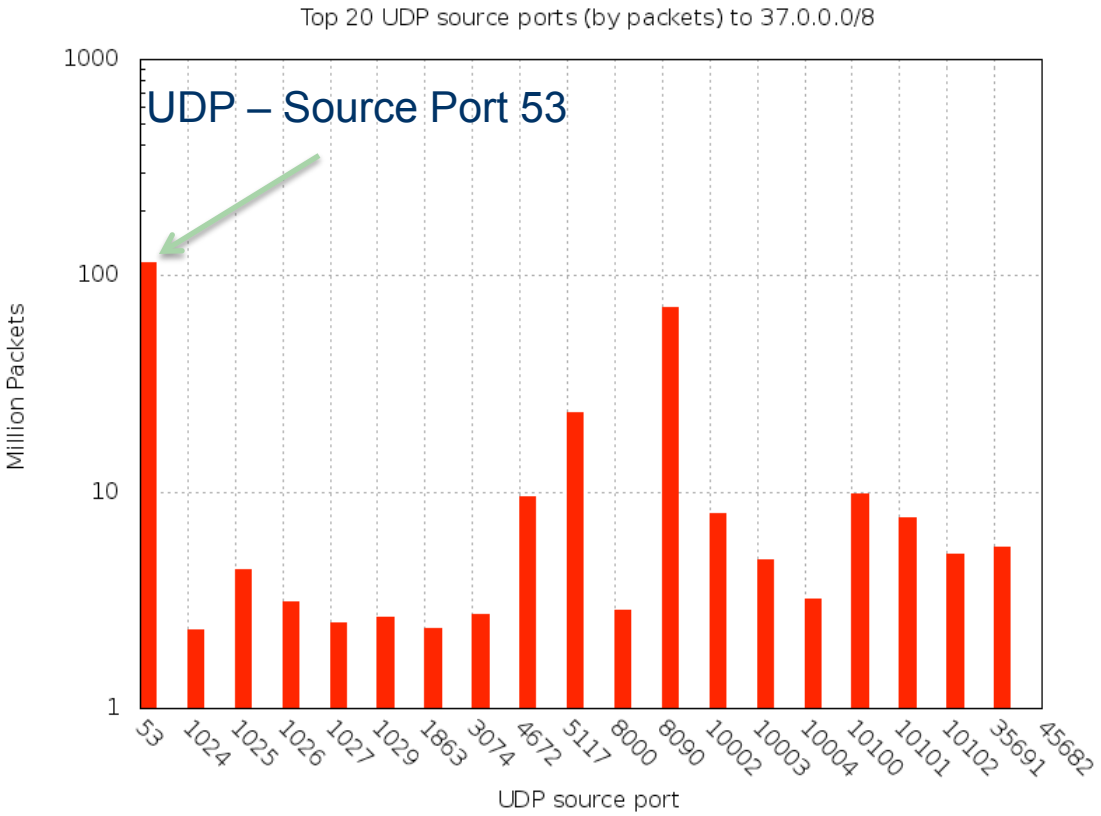




Single /16 and single /24 account for majority of the captured traffic by byte count

37/8


Top 20 TCP destination ports (by packets) to 37.0.0.0/8
TCP – Dest Port 80, 445
Million Packets
100000
10000
1000
100
10
1
0
22
23
25
80
135
139
443
445
1024
TCP destinati


Top 20 UDP source ports (by packets) to 37.0.0.0/8
UDP – Source Port 53
Million Packets
1000
100
10
1
53
1024
1025
1026
1027
1029
1863
3074
4672
5117
8000
8090
10002
10003
10004
10100
10101
10102
35691
45682
UDP source port

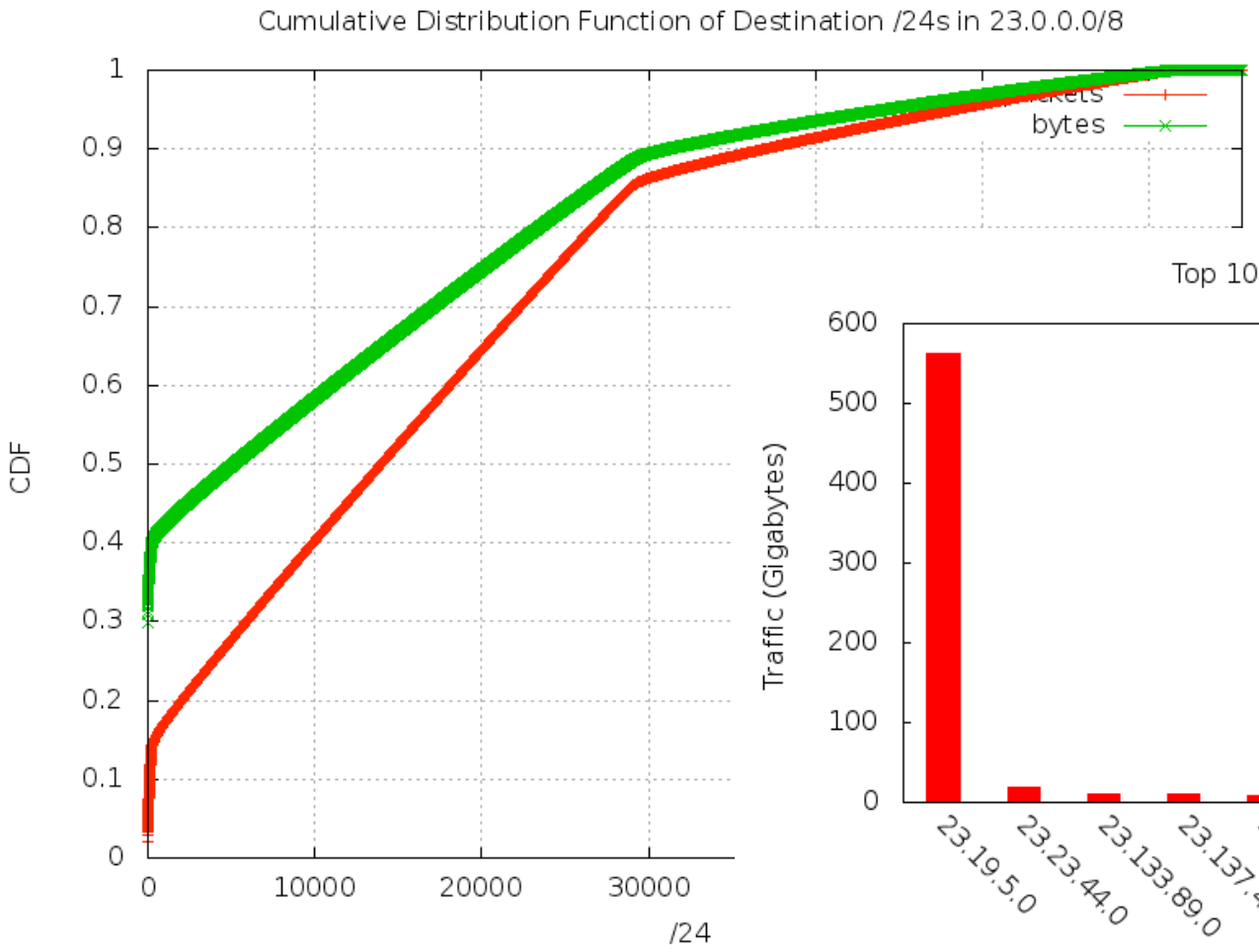
37/8

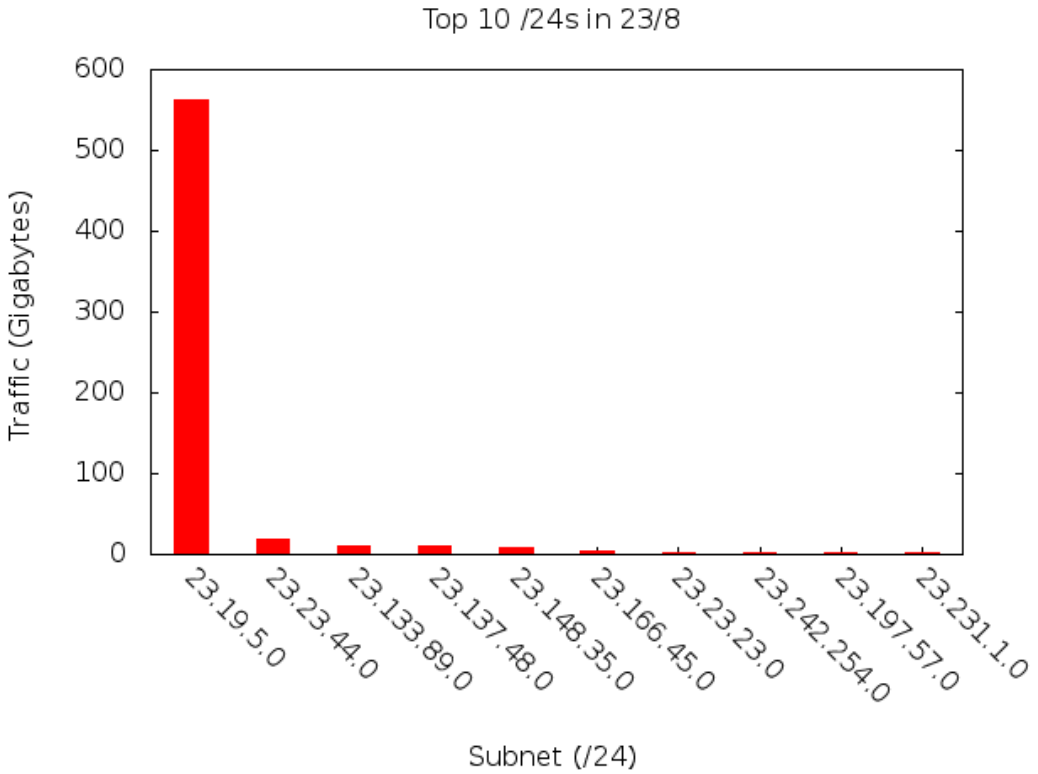
- Port 80 TCP traffic all appears to be directed at single IP address and appears to be related with facebook blocking in china

http://www.renesys.com/blog/2010/06/two-strikes-i-root.shtml

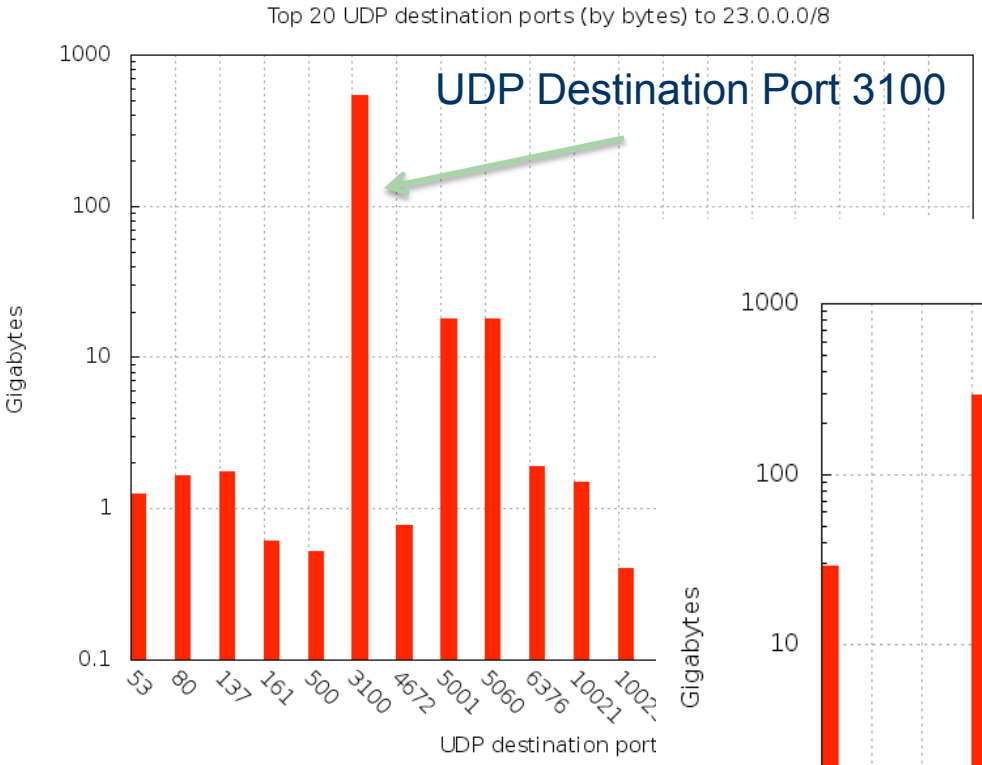
```
dig @dns1.chinatelecom.com.cn. www.facebook.com.
...
   www.facebook.com. 11556 IN A 37.61.54.158
   www.facebook.com. 24055 IN A 78.16.49.15
   www.facebook.com. 38730 IN A 203.98.7.65
```

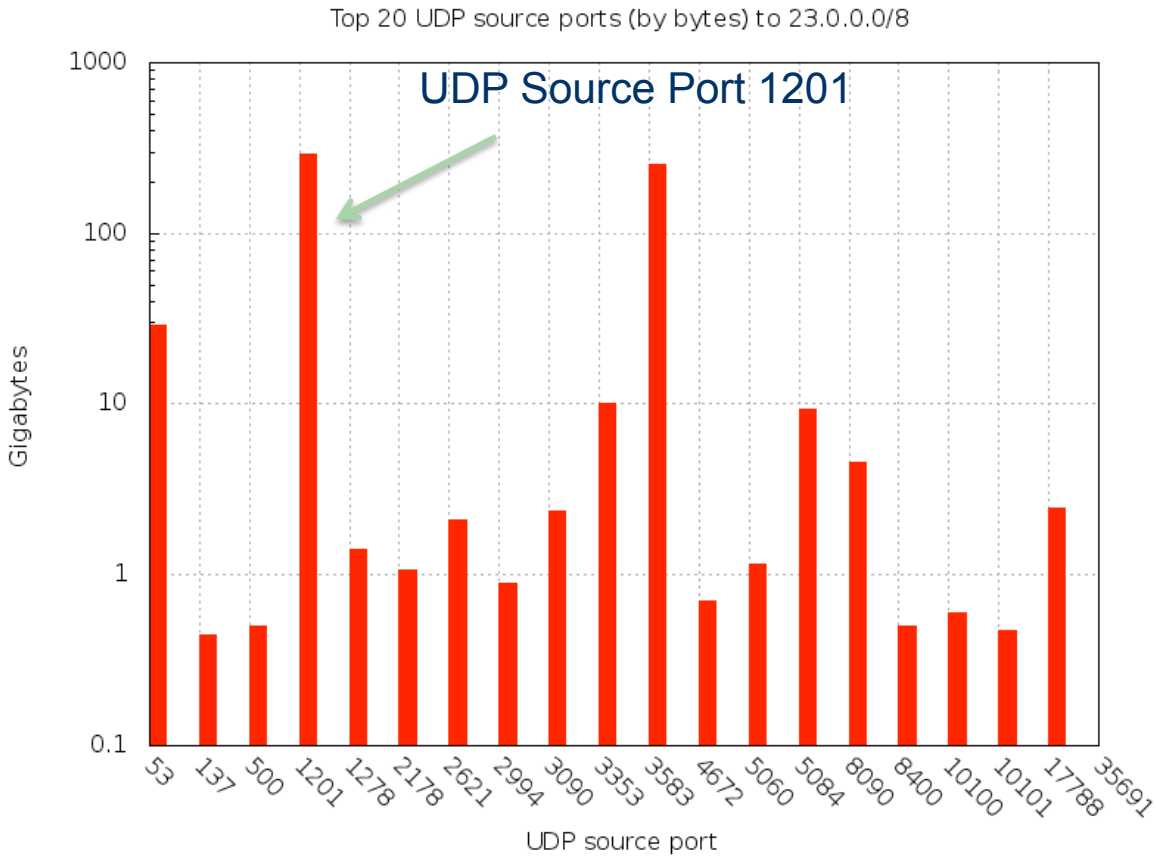
23/8


Cumulative Distribution Function of Destination /24s in 23.0.0.0/8
CDF
/24
bytes
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10000
20000
30000


Top 10 /24s in 23/8
Traffic (Gigabytes)
Subnet (/24)
0
100
200
300
400
500
600
23.19.5.0
23.23.44.0
23.133.89.0
23.137.48.0
23.148.35.0
23.166.45.0
23.23.23.0
23.242.254.0
23.197.57.0
23.231.1.0

23/8

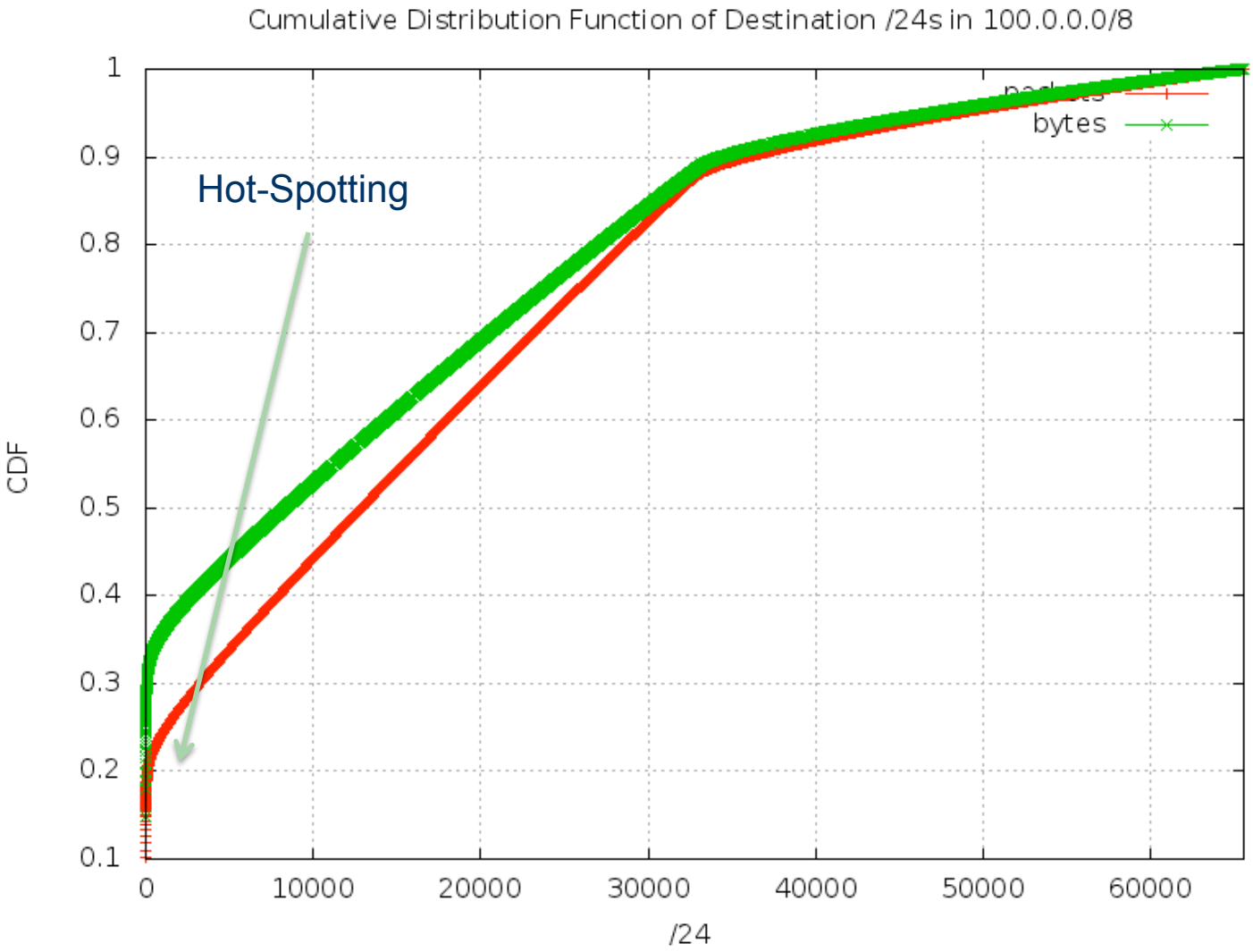

Top 20 UDP destination ports (by bytes) to 23.0.0.0/8
UDP Destination Port 3100
Gigabytes
1000
100
10
1
0.1
53
80
137
161
500
3100
4672
5001
5060
6376
10021
UDP destination port


Top 20 UDP source ports (by bytes) to 23.0.0.0/8
UDP Source Port 1201
Gigabytes
1000
100
10
1
0.1
53
137
500
1201
1278
2178
2621
2994
3090
3353
3583
4672
5060
5084
8090
8400
10100
10101
17788
35691
UDP source port

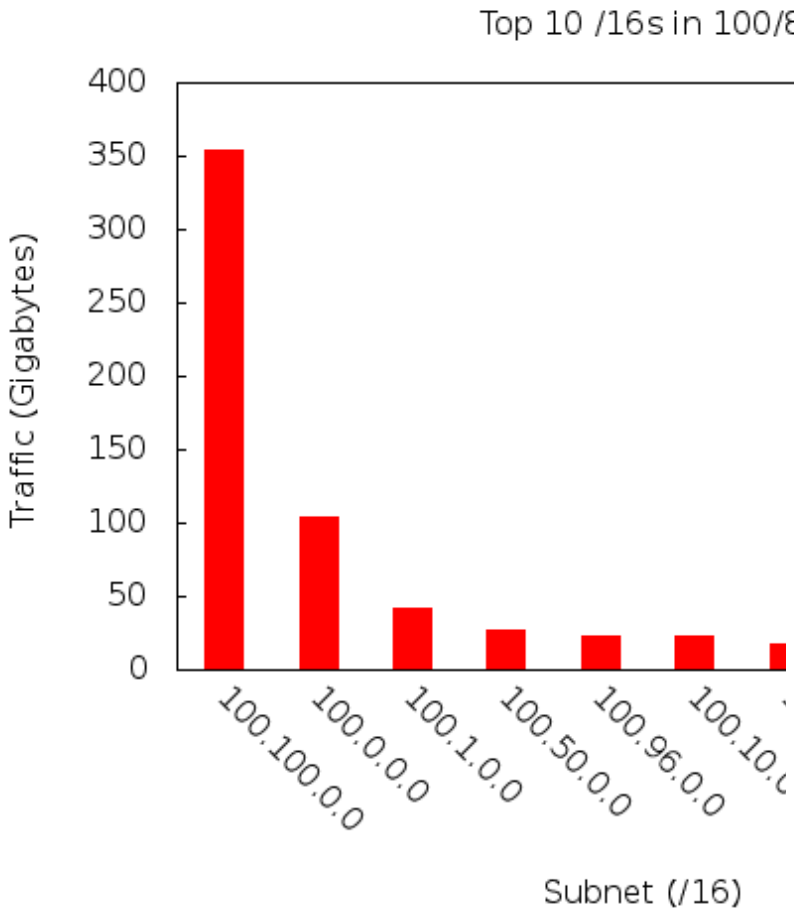
23/8

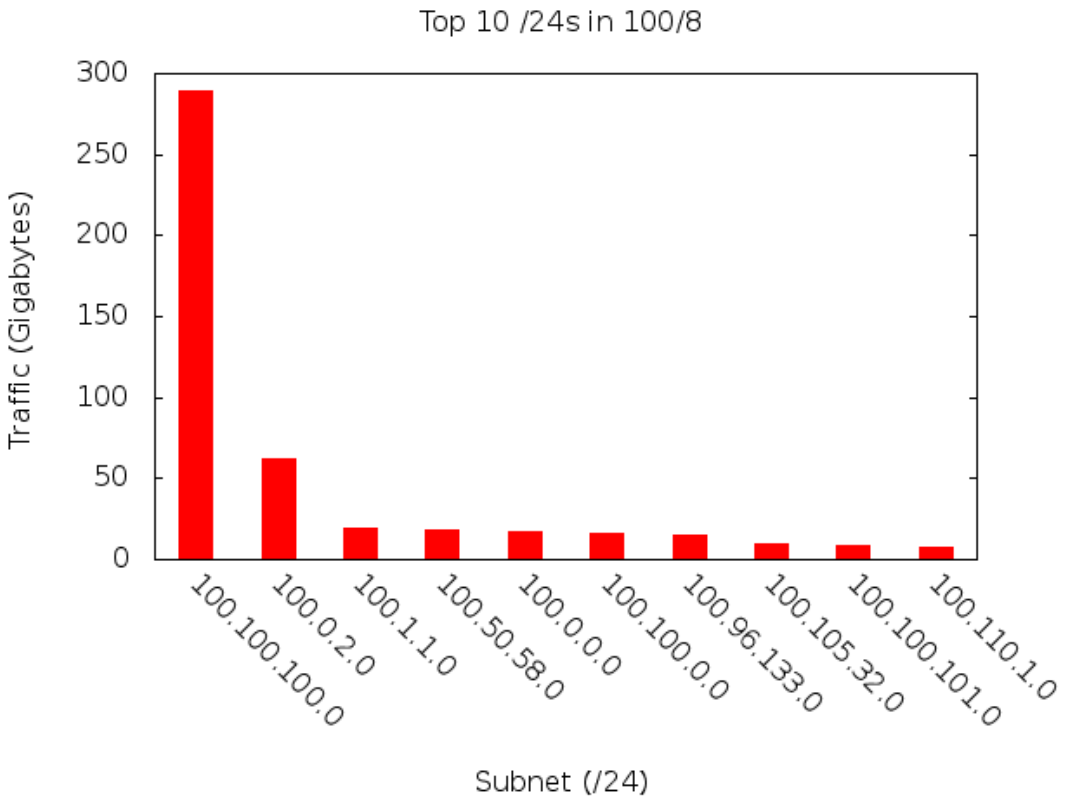
- Traffic from source port 1201 to dest port 3100 from single source to single destination
- Video stream – decoded to 1080p video!

100/8

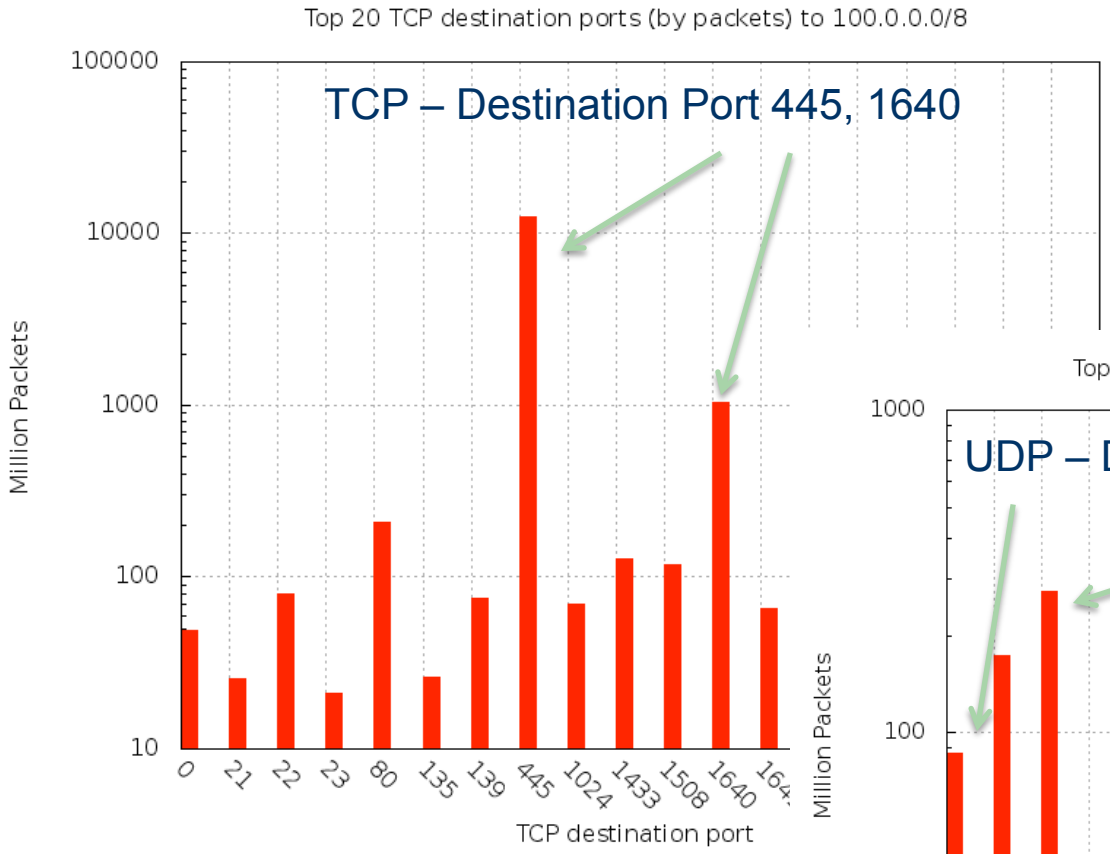

Cumulative Distribution Function of Destination /24s in 100.0.0.0/8
Hot-Spotting
bytes
CDF
/24
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10000
20000
30000
40000
50000
60000

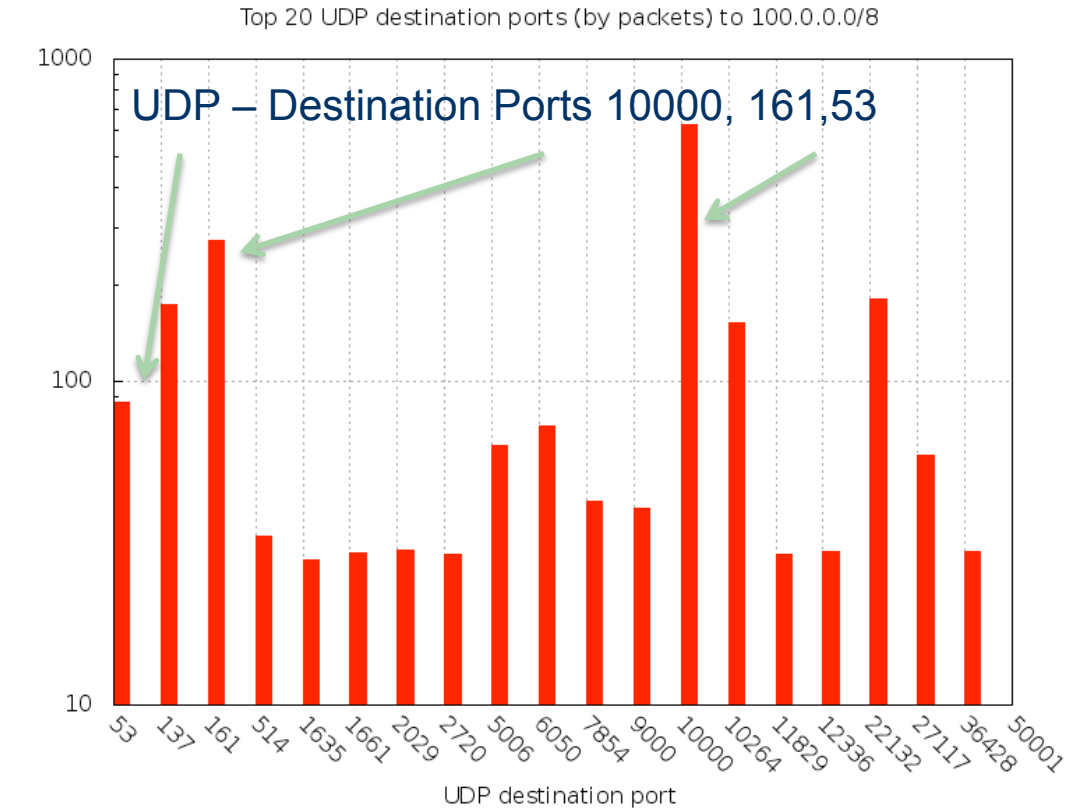
100/8


Top 10 /16s in 100/8
Traffic (Gigabytes)
400
350
300
250
200
150
100
50
0
100.100.0.0
100.0.0.0
100.1.0.0
100.50.0.0
100.96.0.0
Subnet (/16)


Top 10 /24s in 100/8
Traffic (Gigabytes)
300
250
200
150
100
50
0
100.100.100.0
100.0.2.0
100.1.1.0
100.50.58.0
100.0.0.0
100.100.0.0
100.96.133.0
100.105.32.0
100.100.101.0
100.110.1.0
Subnet (/24)

100/8

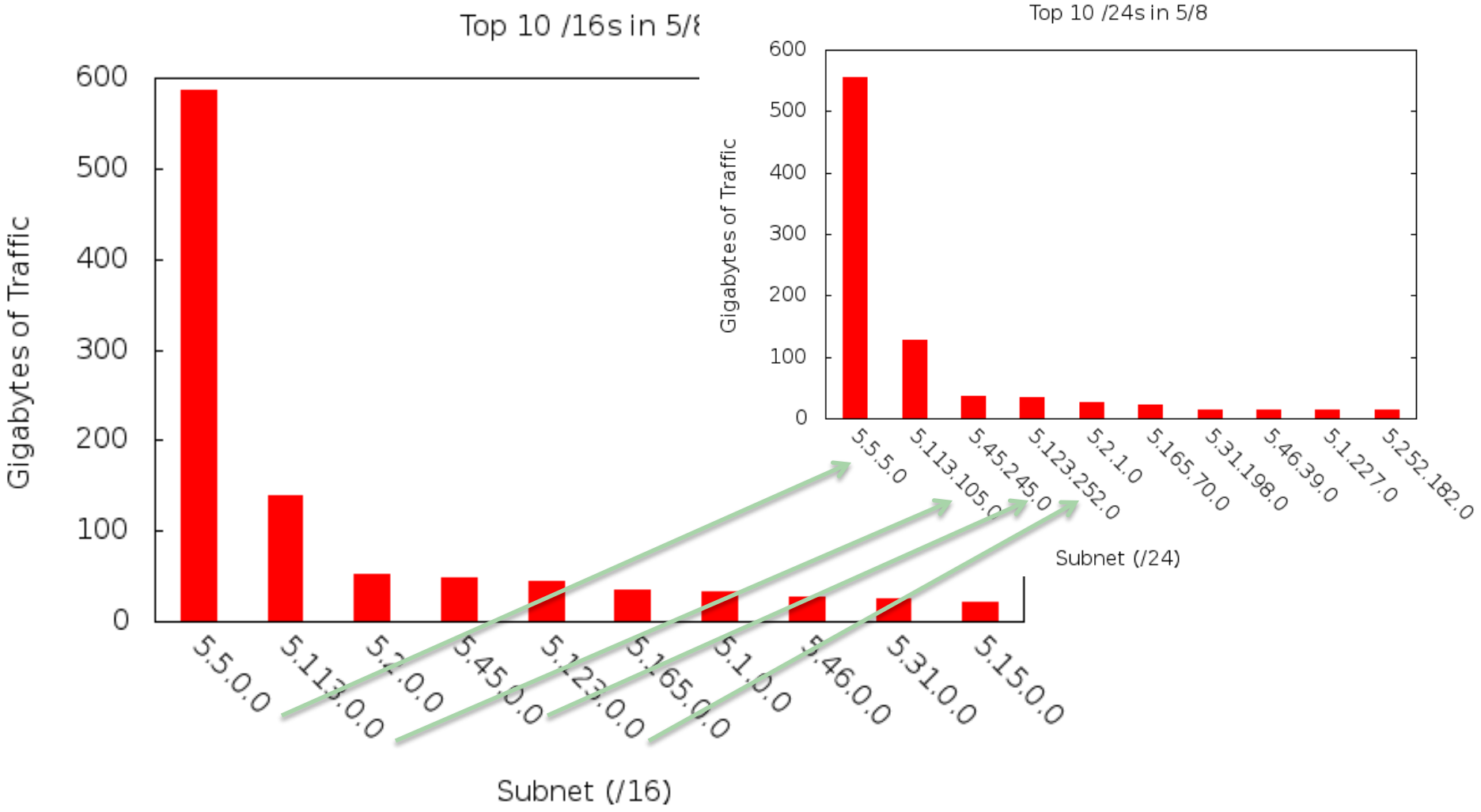

Top 20 TCP destination ports (by packets) to 100.0.0.0/8
TCP – Destination Port 445, 1640
Million Packets
100000
10000
1000
100
10
0
21
22
23
80
135
139
445
1024
1433
1508
1640
TCP destination port


Top 20 UDP destination ports (by packets) to 100.0.0.0/8
UDP – Destination Ports 10000, 161,53
Million Packets
1000
100
10
53
137
161
514
1635
1661
2029
2720
5006
6050
7854
9000
10000
10264
11829
12336
22132
27117
36428
50001
UDP destination port

100/8

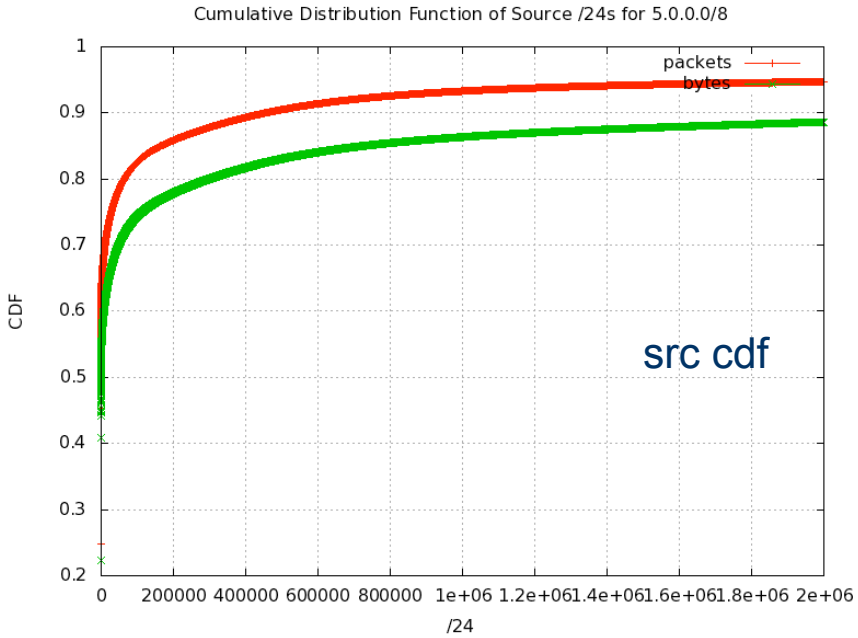
- UDP port 161 – SNMP traffic – default settings in manuals
- UDP source port 10000 – 33 byte packets – micro-torrent some SNMP etc.

5/8

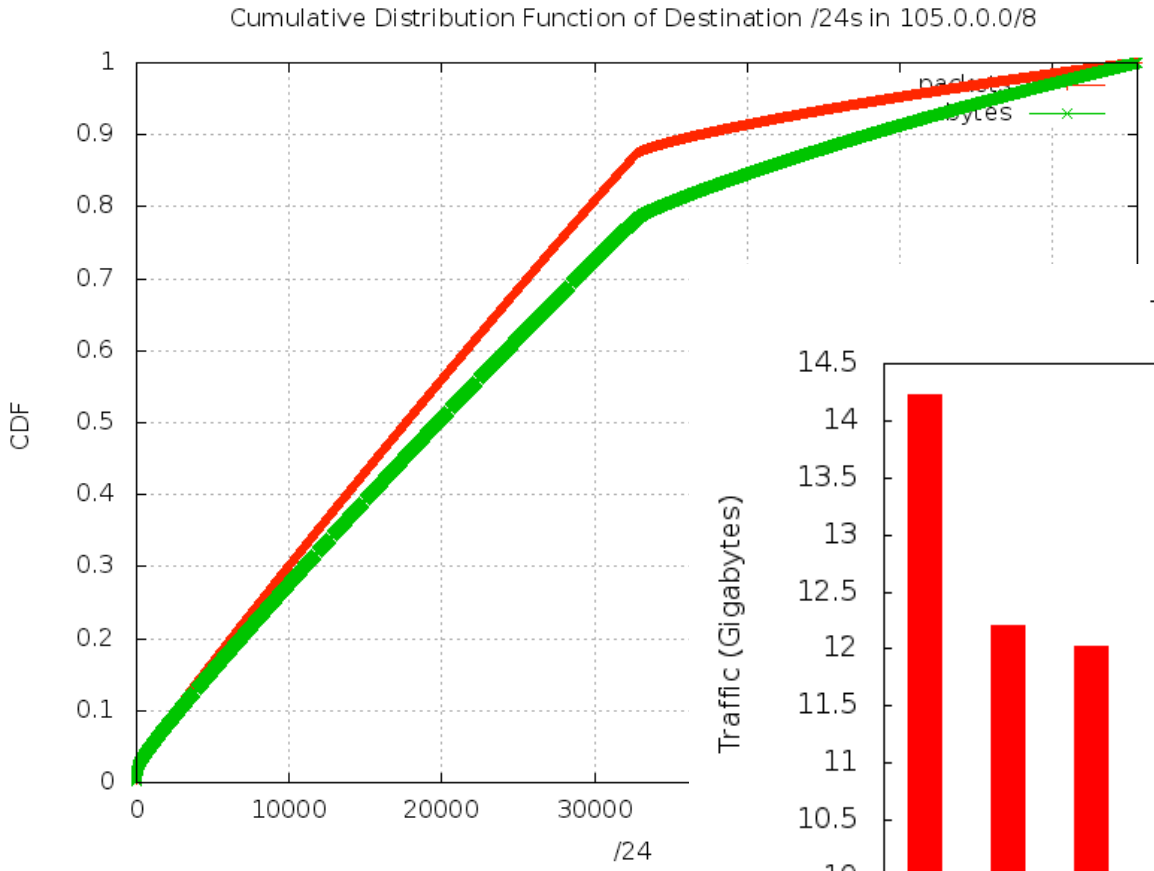

Top 10 /16s in 5/8
Gigabytes of Traffic
600
500
400
300
200
100
0
5.5.0.0
5.113.0.0
5.2.0.0
5.45.0.0
5.123.0.0
5.165.0.0
5.1.0.0
5.46.0.0
5.31.0.0
5.15.0.0
Subnet (/16)
Top 10 /24s in 5/8
Gigabytes of Traffic
600
500
400
300
200
100
0
5.5.5.0
5.113.105.0
5.45.245.0
5.123.252.0
5.2.1.0
5.165.70.0
5.31.198.0
5.46.39.0
5.1.227.0
5.252.182.0
Subnet (/24)

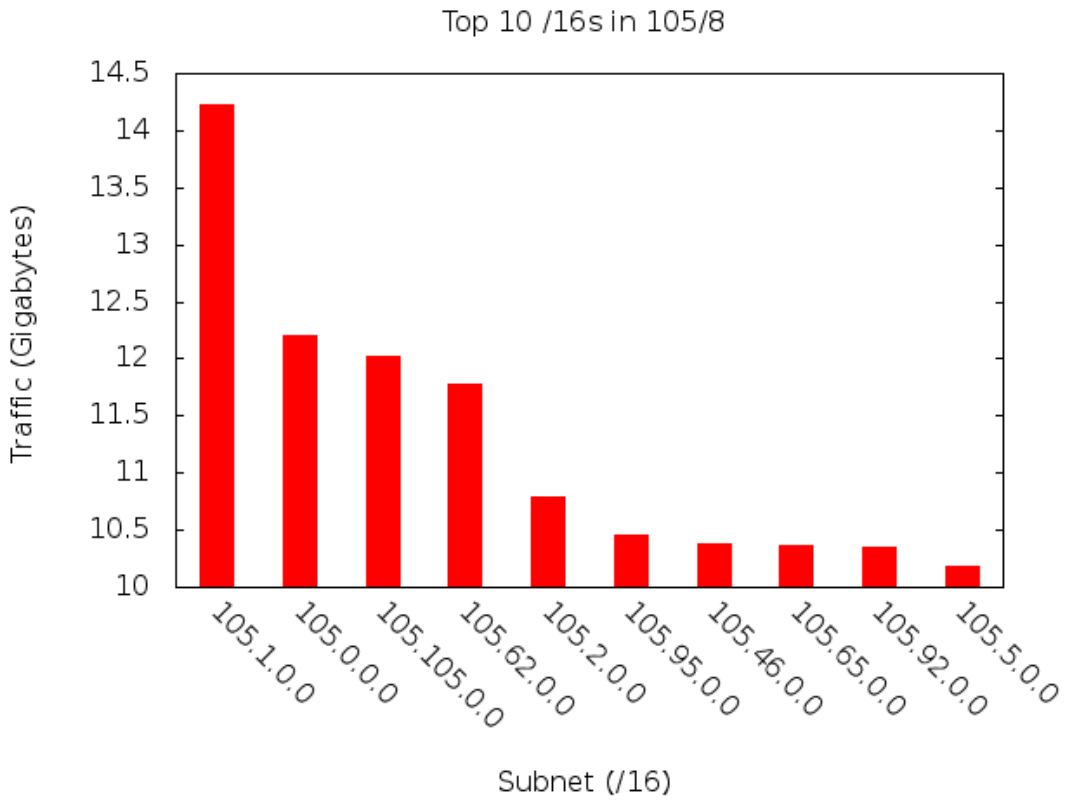
5/8

- Traffic spikes of upto 250Mbps
  - 5.5.5.5 – UDP – 250Byte pkts random ports/srcip
- ICMP6! 5.113.105.0/24
- Flash video – 5.45.245.0/24

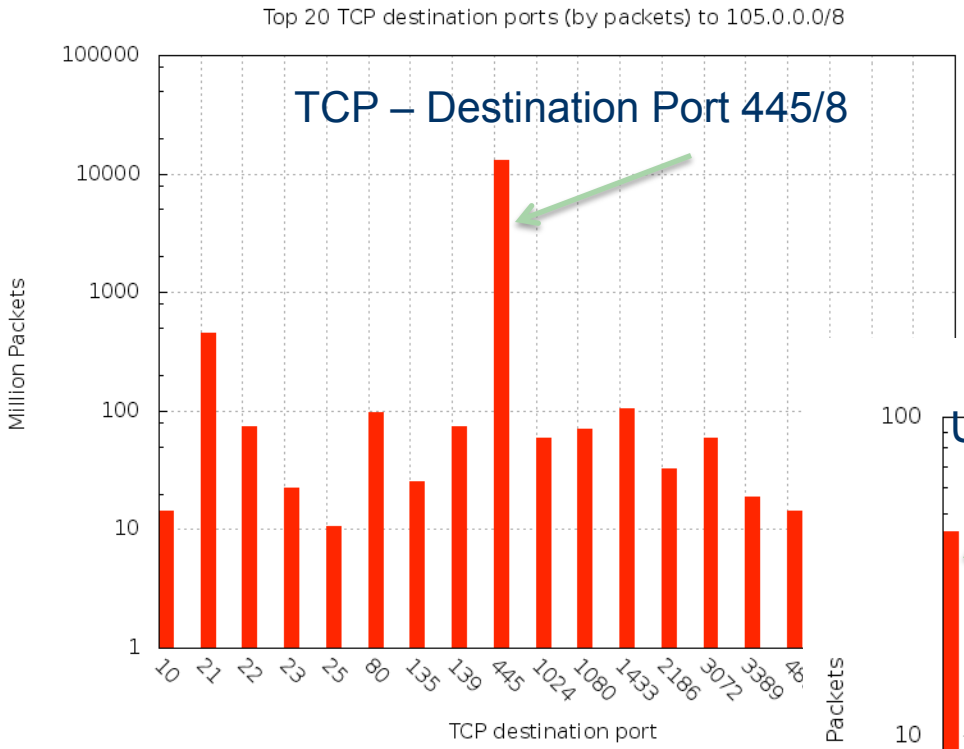


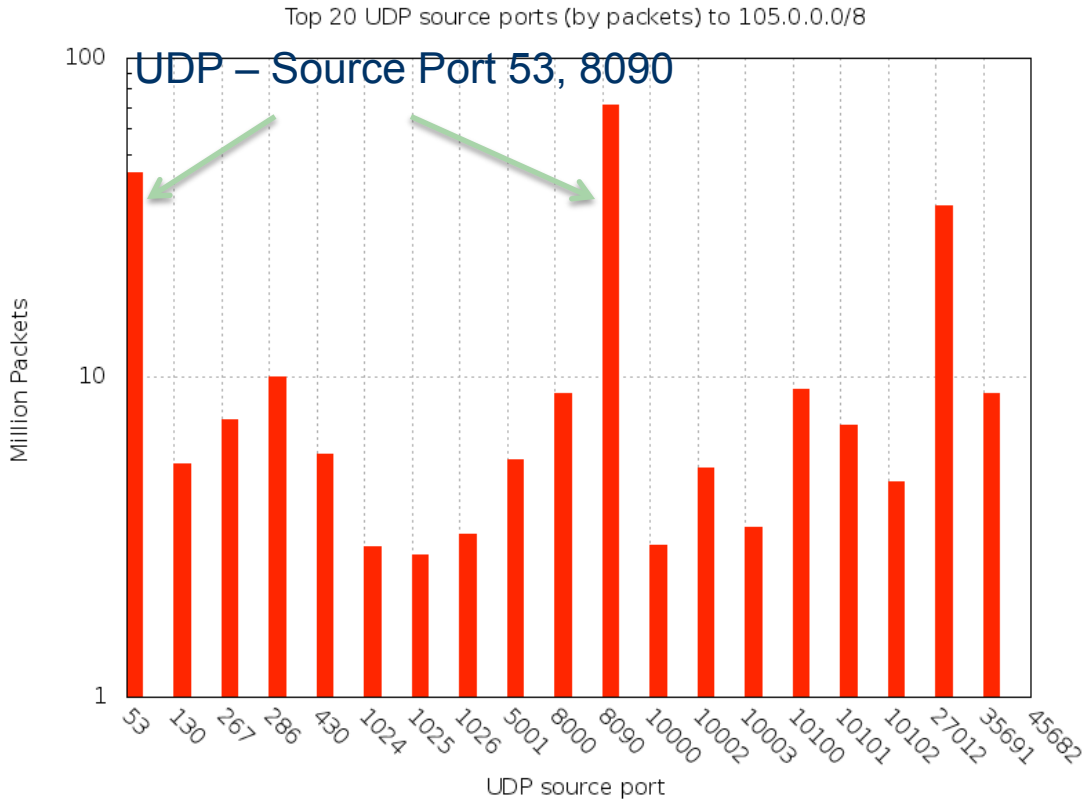
105/8


Cumulative Distribution Function of Destination /24s in 105.0.0.0/8
CDF
/24
bytes
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10000
20000
30000

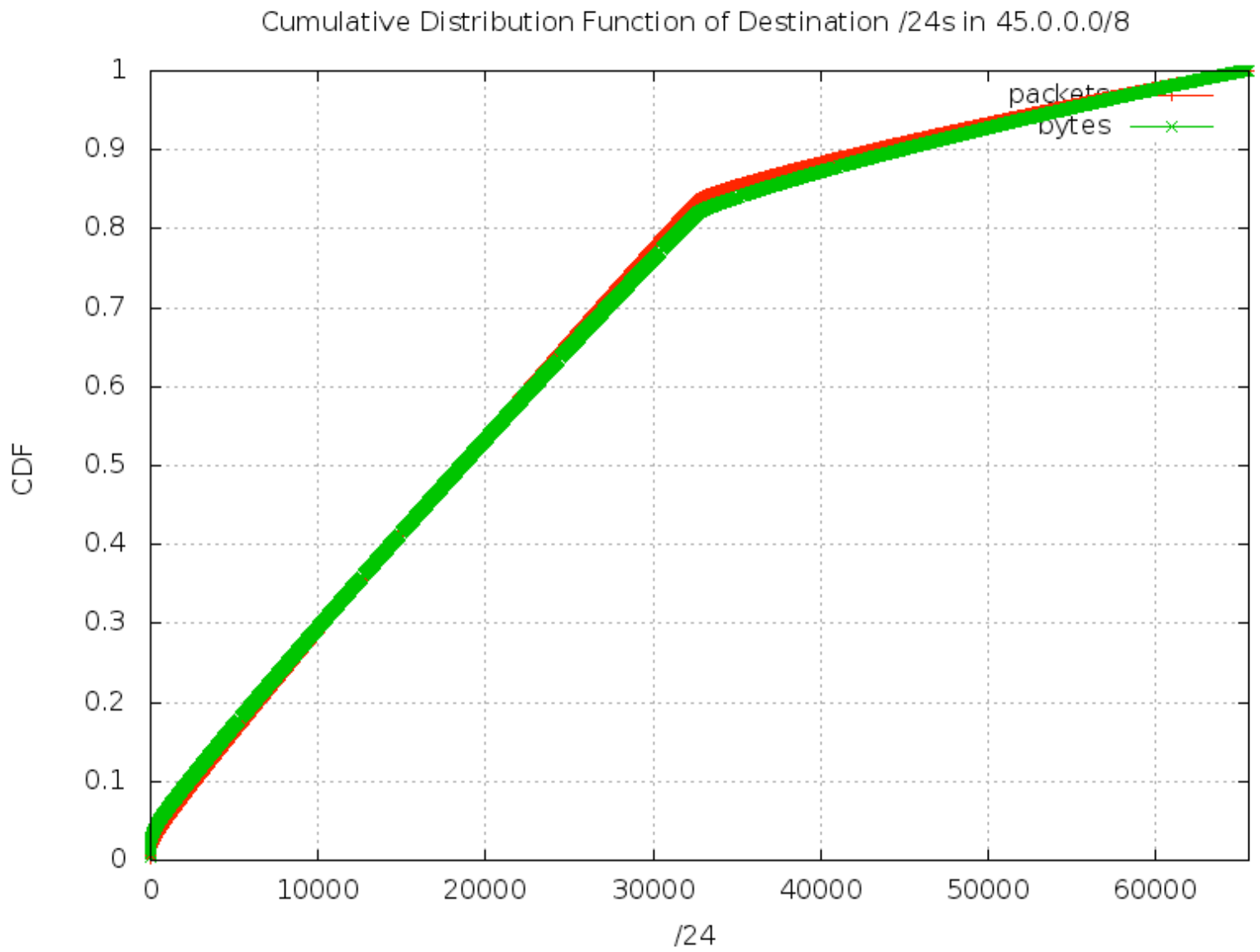

Top 10 /16s in 105/8
Traffic (Gigabytes)
Subnet (/16)
14.5
14
13.5
13
12.5
12
11.5
11
10.5
10
105.1.0.0
105.0.0.0
105.105.0.0
105.62.0.0
105.2.0.0
105.95.0.0
105.46.0.0
105.65.0.0
105.92.0.0
105.5.0.0

105/8

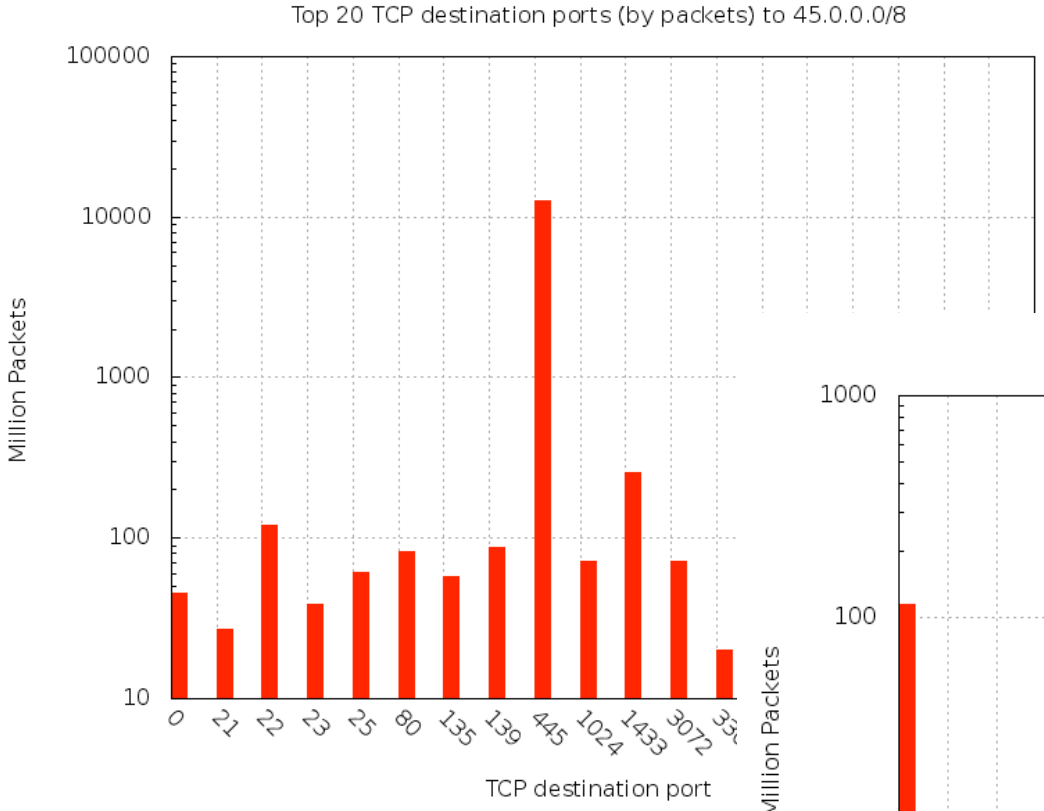

Top 20 TCP destination ports (by packets) to 105.0.0.0/8
TCP – Destination Port 445/8
Million Packets
100000
10000
1000
100
10
1
10
21
22
23
25
80
135
139
445
1024
1080
1433
2186
3072
3389
TCP destination port

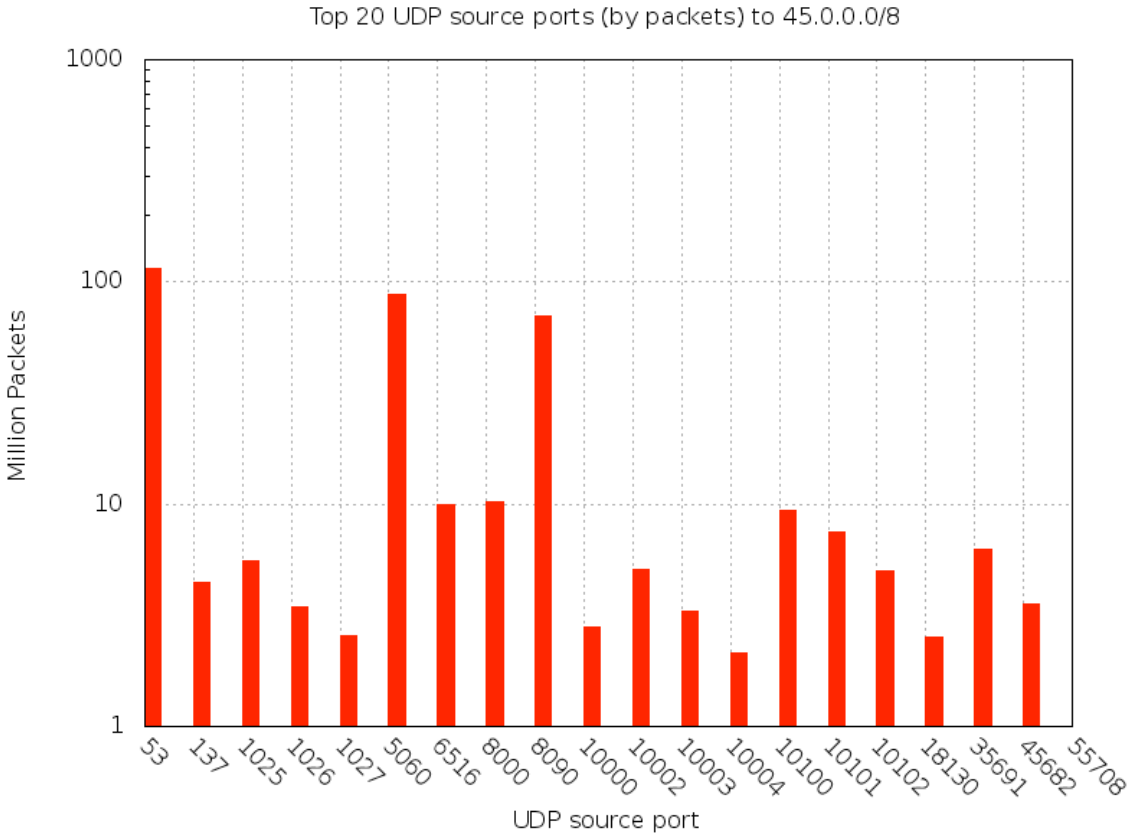

Top 20 UDP source ports (by packets) to 105.0.0.0/8
UDP – Source Port 53, 8090
Million Packets
100
10
1
53
130
267
286
430
1024
1025
1026
5001
8000
8090
10000
10002
10003
10100
10101
10102
27012
35691
45682
UDP source port

45/8

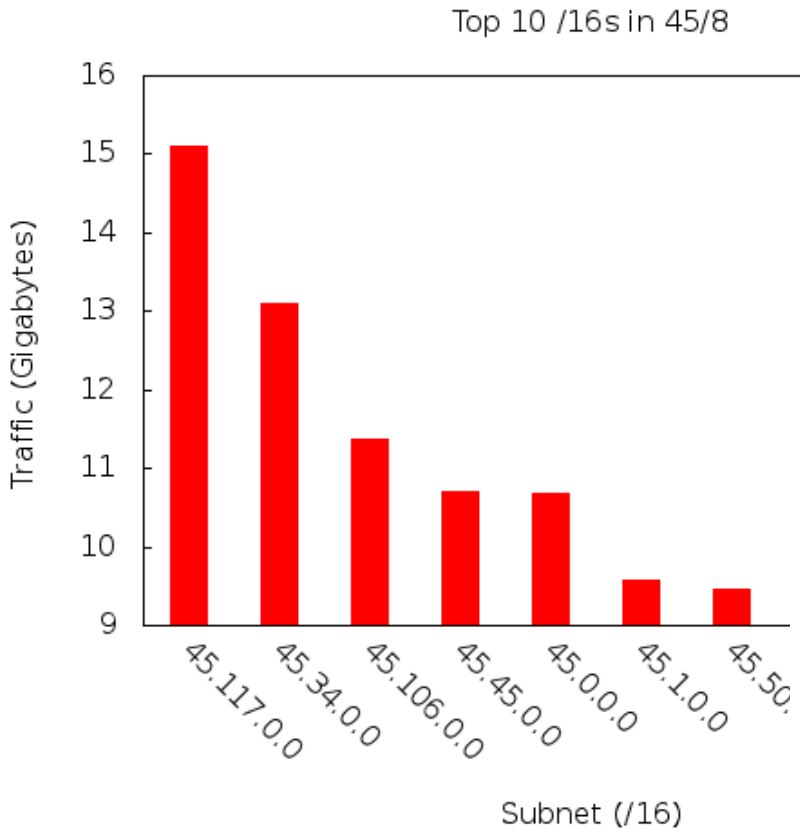

Cumulative Distribution Function of Destination /24s in 45.0.0.0/8
CDF
/24
packets
bytes
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10000
20000
30000
40000
50000
60000

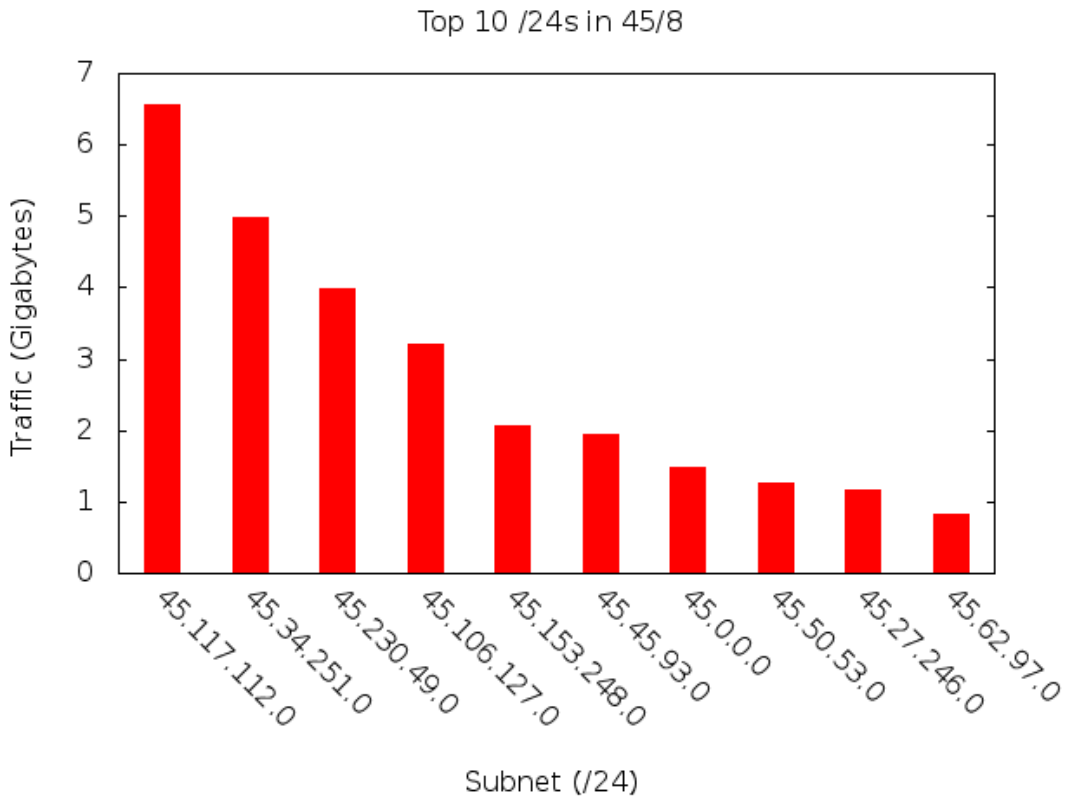
45/8


Top 20 TCP destination ports (by packets) to 45.0.0.0/8
Million Packets
100000
10000
1000
100
10
0
21
22
23
25
80
135
139
445
1024
1433
3072
TCP destination port


Top 20 UDP source ports (by packets) to 45.0.0.0/8
Million Packets
1000
100
10
1
53
137
1025
1026
1027
5060
6516
8000
8090
10000
10002
10003
10004
10100
10101
10102
18130
35691
45682
55708
UDP source port

45/8


Top 10 /16s in 45/8
Traffic (Gigabytes)
16
15
14
13
12
11
10
9
45.117.0.0
45.34.0.0
45.106.0.0
45.45.0.0
45.0.0.0
45.1.0.0
45.50.
Subnet (/16)


Top 10 /24s in 45/8
Traffic (Gigabytes)
7
6
5
4
3
2
1
0
45.117.112.0
45.34.251.0
45.230.49.0
45.106.127.0
45.153.248.0
45.45.93.0
45.0.0.0
45.50.53.0
45.27.246.0
45.62.97.0
Subnet (/24)

Conclusions

- Pollution tends to greatly skew darknet traffic
  - Diverse darknets – diverse reasons for pollution
- General characteristics of background radiation:
  - 15-30Mbps of base traffic for /8 spikes upto 70Mbps
  - Heavily dominated by conficker for TCP traffic
  - DNS for UDP traffic from small set of servers (*)
- Sharing results with relevant RIRs so that they can determine appropriate action regarding cleanup/quarantine

Conclusions

- 100/8, 5/8, 23/8 show relatively abnormal amounts of traffic to portions of the address space
  - Special consideration for:
    - 37.61.54.0/24
    - 23.19.5/24
    - 100.100.100.0/24, 100.0.2.0/24, 100.1.1.0/24
    - 5.5.5.0/24, 5.13.105/24, 5.45.245.0/24, 5.123.252.0/24
  - 45/8, 105/8 relatively clean

Virtual Center for Network and Security Data

Questions?

PREDICT PI MTG

Monday, March 21, 2011

University of California San Diego